1. What are the categorical variables in this dataset? *(depression.csv)*

2. What are the quantitative variables in this dataset? *(depression.csv)*

3. Describe the distribution of the variable "friends" in dataset - Survey that asked 1,200 U.S. college students about their body perception. *(friends.csv)*

4. Describe the distribution of the ages of the Best Actor Oscar winners. Be sure to address shape, center, spread and outliers (Dataset - Best Actor Oscar winners (1970-2013)) *(actor\_age.csv)*

5. Getting information from the output: a. How many observations are in this data set? b. What is the mean age of the actors who won the Oscar? c. What is the five-number summary of the distribution? (Dataset - Best Actor Oscar winners (1970-2013)) *(actor\_age.csv)*

6. Get information from the five-number summary: a. Half of the actors won the Oscar before what age? b. What is the range covered by all the actors' ages? c. What is the range covered by the middle 50% of the ages? (Dataset - Best Actor Oscar winners (1970-2013)) *(actor\_age.csv)*

7. What are the standard deviations of the three rating distributions? Was your intuition correct? (Dataset - 27 students in the class were asked to rate the instructor on a number scale of 1 to 9) *(grad\_data.csv)*

8. Assume that the average rating in each of the three classes is 5 (which should be visually reasonably clear from the histograms), and recall the interpretation of the SD as a "typical" or "average" distance between the data points and their mean. Judging from the table and the histograms, which class would have the largest standard deviation, and which one would have the smallest standard deviation? Explain your reasoning (Dataset - 27 students in the class were asked to rate the instructor on a number scale of 1 to 9) *(ratings.csv)*

Answers:

1.

The categorical variables in this data set are as follows:

1. Hospt
2. Treat
3. Outcome
4. Gender

2.

The quantitative variables in this data set are as follows:

1. Time
2. Age
3. ActuteT

3.

In the given dataset, the distribution of students is not equal. More than half of the students (50.16%) find no difference in their ability to make friends in terms of sex. 36.16% students find it easier to make friends of the opposite sex and 13.66% find it easier to make friends of the same sex.

4.

* The shape of the distribution is skewed right. This means that most actors receive the award at a relatively younger age.
* The center of the the distributions seems to be at around 42-43.
* The data set has a total spread or range from 29 to 76.
* The outlier in the given dataset is of actor of age 76.

5.

1. There are 44 observations in the dataset.
2. Mean age = 44.977
3. 5 number summary is as follows:
   1. Min = 29
   2. Q1 = 38
   3. M = 43.5
   4. Q3 = 50.5
   5. Max = 76

6.

1. 43.5 (Median)
2. Range varies from 29 – 76
3. Range covered by the middle 50% is 38 – 50.5

7.

Standard deviations of the three classes are as follows:

1. Class 1: 1.56892908
2. Class 2: 4
3. Class 3: 2.63117406

8. Class 1 has the smallest standard deviation, because most of the values are close to the mean (5).

Class 2 has the greatest standard deviation, since most of the values are distant from the mean (5).